CLAIMS

What is claimed is:

1   1. A method for processing a speech signal, comprising:

2   receiving an input speech signal;

3   constructing a phoneme lattice for the input speech signal;

4   searching the phoneme lattice to produce a likelihood score for each

5   potential path; and

6   determining a processing result for the input speech signal based on the

7   likelihood score of each potential path.

8

1   2. The method of claim 1, wherein constructing the phoneme lattice

2   comprises:

3   segmenting an input speech signal into frames;

4   extracting acoustic features for a frame of the input speech signal;

5   determining K-best initial phoneme paths leading to the frame based on a

6   first score of each potential phoneme path leading to the frame; and

7   calculating a second score for each of the K-best phoneme paths for the

8   frame.

9

1   3. The method of claim 2, further comprising:

2   clustering together K-best initial phoneme paths for at least one

3   consecutive frame;

4   selecting M-best refined phoneme paths among the clustered phoneme

5   paths based on second scores of these paths; and

6   identifying vertices and arc parameters of the phoneme lattice for the

7   input speech signal.

8

1   4. The method of claim 2, wherein the first score and the second score

2   comprise a score based on phoneme acoustic models and language models.

3

1   5. The method of claim 1, wherein searching the phoneme lattice
2   comprises:
3   receiving a phoneme lattice;
4   traversing the phoneme lattice via potential paths;
5   computing a score for a traversed path based on at least one of a
6   phoneme confusion matrix and a plurality of language models; and
7   modifying the score for the traversed path.
8

1   6. The method of claim 5, wherein modifying the score comprises
2   adjusting the score by at least one of the following: allowing repetition of
3   phonemes and allowing flexible endpoints for phonemes in a path.
4

1   7. The method of claim 1, wherein determining the processing result
2   comprises determining at least one of the following: at least one candidate
3   textual representation of the input speech signal and a likelihood that the input
4   speech signal contains targeted keywords.
5

1   8. A method for constructing a phoneme lattice for an input audio signal
2   comprising:
3   segmenting the input audio signal into frames;
4   extracting acoustic features for a frame of the input audio signal;
5   determining K-best initial phoneme paths leading to the frame based on a
6   first score of each potential phoneme path leading to the frame; and
7   calculating a second score for each of the K-best phoneme paths for the
8   frame.
9

1   9. The method of claim 8, further comprising:
2   clustering together K-best initial phoneme paths for at least one
3   consecutive frame;
4   selecting M-best refined phoneme paths among the clustered phoneme
5   paths based on second scores of these paths; and

6   identifying vertices and arc parameters of the phoneme lattice for the

7   input speech signal.

8

1   10. The method of claim 8, wherein the first score and the second score

2   comprises a score based on phoneme acoustic models and language models.

3

1   11. A method for searching a phoneme lattice, comprises:

2   receiving a phoneme lattice;

3   traversing the phoneme lattice via potential paths; and

4   computing a score for a traversed path based on at least one of a

5   phoneme confusion matrix and a plurality of language models.

6

1   12. The method of claim 11, further comprising modifying the score for

2   the traversed path.

3

1   13. The method of claim 12, wherein modifying the score comprises

2   adjusting the score by at least one of the following: allowing repetition of

3   phonemes and allowing flexible endpoints for phonemes in a path.

4

1   14. The method of claim 11, further comprising determining a search

2   result for the input audio signal based on the modified score of each searched

3   path.

4

1   15. A method for distributing speech processing, comprising:

2   receiving an input speech signal by a client;

3   constructing a phoneme lattice for the input speech signal by the client;

4   transmitting the phoneme lattice from the client to a server; and

5   searching the phoneme lattice to produce a result for the input speech

6   signal for the purpose of at least one of recognizing speech and spotting

7   keywords, in the input speech signal.

8

1   16. The method of claim 15, wherein constructing the phoneme lattice

2 comprises:

3   segmenting an input speech signal into frames;

4   extracting acoustic features for a frame of the input speech signal;

5   determining K-best initial phoneme paths leading to the frame based on a

6 first score of each potential phoneme path leading to the frame; and

7   calculating a second score for each of the K-best phoneme paths.

8

1   17. The method of claim 16, further comprising:

2   clustering together K-best initial phoneme paths for at least one

3 consecutive frame;

4   selecting M-best refined phoneme paths among the clustered phoneme

5 paths based on second scores of these paths; and

6   identifying vertices and arc parameters of the phoneme lattice for the

7 input speech signal.

8

1   18. The method of claim 16, wherein the first score and the second score

2 comprise a score based on phoneme acoustic models and phoneme language

3 models.

4

1   19. The method for claim 15, wherein searching the phoneme lattice

2 comprises:

3   receiving a phoneme lattice;

4   traversing the phoneme lattice via potential paths;

5   computing a likelihood score for a traversed path based on at least a

6 phoneme confusion matrix and a plurality of language models;

7   modifying the score for the traversed path; and

8   determining a search result for the input audio signal based on the

9 modified score of each searched path.

10

1   20. The method of claim 19, wherein modifying the score comprises

2   adjusting the score by at least one of the following: allowing repetition of

3   phonemes and allowing flexible endpoints for phonemes in a path.

4

1   21. A method for training a phoneme confusion matrix, comprising:

2   initializing the phoneme confusion matrix;

3   estimating confusion probabilities between phonemes based on a training

4   database, and the initial phoneme confusion matrix; and

5   updating the phoneme confusion matrix based on the estimated confusion

6   probabilities.

7

1   22. The method of claim 21, wherein the training database comprises a

2   plurality of utterances, actual phoneme sequences corresponding to the plurality

3   of utterances, and time alignment information between utterances and actual

4   phoneme sequences of the utterances.

5

1   23. The method of claim 21, wherein estimating the confusion

2   probabilities comprises:

3   constructing a phoneme lattice for each utterance in the training

4   database;

5   searching the phoneme lattice to produce a phoneme sequence

6   hypothesis for the corresponding utterance; and

7   estimating the confusion probabilities between phonemes based on

8   statistics obtained by comparing actual phoneme sequences and corresponding

9   phoneme sequence hypotheses.

10

1   24. A speech processing system, comprising:

2   a phoneme lattice constructor to construct a phoneme lattice for an input

3   speech signal;

4   a phoneme lattice search mechanism to search the phoneme lattice for

5   the purpose of at least of recognizing speech and spotting keywords, in the input

6   speech signal;

7   a plurality of models for lattice construction; and

8   a plurality of models for lattice search.

9

1   25. The system of claim 24, wherein the phoneme lattice constructor

2   comprises:

3   an acoustic feature extractor to segment the input speech signal into

4   frames and to extract acoustic features for a frame;

5   a phoneme path estimator to determine K-best initial phoneme paths

6   leading to the frame;

7   a global score evaluator to determine M-best refined phoneme paths

8   based on a cluster of K-best paths of at least one consecutive frame; and

9   a lattice parameter identifier to identify lattice vertices and arc parameters

10   based on M-best refined phoneme paths of each frame.

11

1   26. The system of claim 24, wherein the plurality of models for lattice

2   construction comprise a plurality of phoneme acoustic models and a plurality of

3   language models.

4

1   27. The system of claim 24, wherein the plurality of models for lattice

2   search comprise a phoneme confusion matrix and a plurality of language

3   models.

4

1   28. A system for constructing a phoneme lattice, comprising:

2   an acoustic feature extractor to segment an input speech signal into

3   frames and to extract acoustic features for a frame;

4   a phoneme path estimator to determine K-best initial phoneme paths

5   leading to the frame;

6   a global score evaluator to determine M-best refined phoneme paths

7   based on a cluster of K-best paths of at least one consecutive frame; and

8   a lattice parameter identifier to identify lattice vertices and arc parameters

9   based on M-best refined phoneme paths of each frame.

10

1   29. The system of claim 28, wherein the phoneme path estimator

2   comprises a likelihood score evaluator to calculate a first score for a potential

3   phoneme path leading to each frame.

4

1   30. The system of claim 28, wherein the global score evaluator comprises

2   a score computation component to calculate a second score for each of K-best

3   initial phoneme paths for each frame.

4

1   31. A distributed speech processing system, comprising:

2   a client to receive an input speech signal and to construct a phoneme

3   lattice for the input speech signal; and

4   a server to search the phoneme lattice to produce a result for the input

5   speech signal for the purpose of at least one of recognizing speech and spotting

6   keywords, in the input speech signal.

7

1   32. The system of claim 31, wherein the client comprises a phoneme

2   lattice constructor to construct a phoneme lattice and a transmitting component

3   to transmit the phoneme lattice to the server.

4

1   33. The system of claim 31, wherein the server comprises a receiving

2   component to receive the phoneme lattice from the client and a phoneme lattice

3   search mechanism to search the phoneme lattice.

4

1   34. A system for training a phoneme confusion matrix, comprising:

2   a confusion matrix initializer to initialize the phoneme confusion matrix;

3   a phoneme lattice constructor to construct a phoneme lattice for each

4   utterance in a training database; and

5   a phoneme lattice search mechanism to search the phoneme lattice to

6   produce a phoneme sequence hypothesis for the corresponding utterance,

7   based on the initial phoneme confusion matrix and a plurality of language

8   models.

9

1   35. The system of claim 34, further comprising a confusion matrix

2   updater to update the initial phoneme confusion matrix using confusion

3   probabilities between phonemes estimated from statistics obtained by comparing

4   actual phoneme sequences and corresponding phoneme sequence hypotheses.

5

1   36. The system of claim 35, wherein the phoneme confusion matrix

2   updater comprises a confusion probability estimator to estimate confusion

3   probabilities between phonemes based on the training database.

4

1   37. An article comprising: a machine accessible medium having content

2   stored thereon, wherein when the content is accessed by a processor, the

3   content provides for processing a speech signal by:

4   receiving an input speech signal;

5   constructing a phoneme lattice for the input speech signal;

6   searching the phoneme lattice to produce a likelihood score for each

7   potential path; and

8   determining a processing result for the input speech signal based on the

9   likelihood score of each potential path.

10

1   38. The article of claim 37, wherein content for constructing the phoneme

2   lattice comprises content for:

3   segmenting an input speech signal into frames;

4   extracting acoustic features for a frame of the input speech signal;

5   determining K-best initial phoneme paths leading to the frame based on a

6   first score of each potential phoneme path leading to the frame; and

7   calculating a second score for each of the K-best phoneme paths for the

8   frame.

9

1   39. The article of claim 38, further comprising content for:

2   clustering together K-best initial phoneme paths for at least one

3   consecutive frame;

4   selecting M-best refined phoneme paths among the clustered phoneme

5   paths based on second scores of these paths; and

6   identifying vertices and arc parameters of the phoneme lattice for the

7   input speech signal.

8

1   40. The article of claim 38, wherein the first score and the second score

2   comprise a score based on phoneme acoustic models and language models.

3

1   41. The article of claim 37, wherein content for searching the phoneme

2   lattice comprises content for:

3   receiving a phoneme lattice;

4   traversing the phoneme lattice via potential paths;

5   computing a score for a traversed path based on at least one of a

6   phoneme confusion matrix and a plurality of language models; and

7   modifying the score for the traversed path.

8

1   42. The article of claim 41, wherein content for modifying the score

2   comprises content for adjusting the score by at least one of the following:

3   allowing repetition of phonemes and allowing flexible endpoints for phonemes in

4   a path.

5

1   43. The article of claim 37, wherein content for determining the

2   processing result comprises content for determining at least one of the following:

3   at least one candidate textual representation of the input speech signal and a

4   likelihood that the input speech signal contains targeted keywords.

5

1   44. An article comprising: a machine accessible medium having content

2   stored thereon, wherein when the content is accessed by a processor, the

3   content provides for constructing a phoneme lattice for an input audio signal by:

4   segmenting the input audio signal into frames;

5   extracting acoustic features for a frame of the input audio signal;

6   determining K-best initial phoneme paths leading to the frame based on a

7   first score of each potential phoneme path leading to the frame; and

8   calculating a second score for each of the K-best phoneme paths for the

9   frame.

10

1   45. The article of claim 44, further comprising content for:

2   clustering together K-best initial phoneme paths for at least one

3   consecutive frame;

4   selecting M-best refined phoneme paths among the clustered phoneme

5   paths based on second scores of these paths; and

6   identifying vertices and arc parameters of the phoneme lattice for the

7   input speech signal.

8

1   46. The article of claim 44, wherein the first score and the second score

2   comprises a score based on phoneme acoustic models and language models.

3

1   47. An article comprising: a machine accessible medium having content

2   stored thereon, wherein when the content is accessed by a processor, the

3   content provides for searching a phoneme lattice by:

4   receiving a phoneme lattice;

5   traversing the phoneme lattice via potential paths; and

6   computing a score for a traversed path based on at least one of a

7   phoneme confusion matrix and a plurality of language models.

8

1   48. The article of claim 47, further comprising content for modifying the

2   score for the traversed path.

3

1   49. The article of claim 48, wherein content for modifying the score

2   comprises content for adjusting the score by at least one of the following:

3   allowing repetition of phonemes and allowing flexible endpoints for phonemes in

4   a path.

5

1   50. The article of claim 47, further comprising content for determining a

2   search result for the input audio signal based on the modified score of each

3   searched path.

4

1   51. An article comprising: a machine accessible medium having content

2   stored thereon, wherein when the content is accessed by a processor, the

3   content provides for distributing speech processing by:

4   receiving an input speech signal by a client;

5   constructing a phoneme lattice for the input speech signal by the client;

6   transmitting the phoneme lattice from the client to a server; and

7   searching the phoneme lattice to produce a result for the input speech

8   signal for the purpose of at least one of recognizing speech and spotting

9   keywords, in the input speech signal.

10

1   52. The article of claim 51, wherein content for constructing the phoneme

2   lattice comprises content for:

3   segmenting an input speech signal into frames;

4   extracting acoustic features for a frame of the input speech signal;

5   determining K-best initial phoneme paths leading to the frame based on a

6   first score of each potential phoneme path leading to the frame; and

7   calculating a second score for each of the K-best phoneme paths.

8

1   53. The article of claim 52, further comprising content for:

2   clustering together K-best initial phoneme paths for at least one

3   consecutive frame;

4   selecting M-best refined phoneme paths among the clustered phoneme

5   paths based on second scores of these paths; and

6   identifying vertices and arc parameters of the phoneme lattice for the

7   input speech signal.

8

1   54. The article of claim 52, wherein the first score and the second score

2   comprise a score based on phoneme acoustic models and phoneme language

3   models.

4

1   55. The article for claim 51, wherein content for searching the phoneme

2   lattice comprises content for:

3   receiving a phoneme lattice;

4   traversing the phoneme lattice via potential paths;

5   computing a likelihood score for a traversed path based on at least a

6   phoneme confusion matrix and a plurality of language models;

7   modifying the score for the traversed path; and

8   determining a search result for the input audio signal based on the

9   modified score of each searched path.

10

1   56. The article of claim 55, wherein content for modifying the score

2   comprises content for adjusting the score by at least one of the following:

3   allowing repetition of phonemes and allowing flexible endpoints for phonemes in

4   a path.

5

1   57. An article comprising: a machine accessible medium having content

2   stored thereon, wherein when the content is accessed by a processor, the

3   content provides for training a phoneme confusion matrix by:

4   initializing the phoneme confusion matrix;

5   estimating confusion probabilities between phonemes based on a training

6 database, and the initial phoneme confusion matrix; and

7   updating the phoneme confusion matrix based on the estimated confusion

8 probabilities.

9

1   58. The article of claim 57, wherein the training database comprises a

2 plurality of utterances, actual phoneme sequences corresponding to the plurality

3 of utterances, and time alignment information between utterances and actual

4 phoneme sequences of the utterances.

5

1   59. The article of claim 57, wherein content for estimating the confusion

2 probabilities comprises content for:

3   constructing a phoneme lattice for each utterance in the training

4 database;

5   searching the phoneme lattice to produce a phoneme sequence

6 hypothesis for the corresponding utterance; and

7   estimating the confusion probabilities between phonemes based on

8 statistics obtained by comparing actual phoneme sequences and corresponding

9 phoneme sequence hypotheses.

10